PENGEMBANGAN INSTRUMEN KEMAMPUAN BERPIKIR KRITIS SISWA PADA MATERI DAMPAK SOSIAL INFORMATIKA

C. Rizkani[1], Sarwi[2]

[12]Program Studi Penelitian dan Evaluasi Pendidikan
Universitas Negeri Semarang
Semarang, Indonesia

e-mail: rizkanicut2000@students.unnes.ac.id[1] , sarwi_dosen@mail.unnes.ac.id[2]

Abstrak

Kemampuan berpikir kritis merupakan salah satu kompetensi penting dalam pembelajaran Informatika, terutama ketika siswa dihadapkan pada berbagai persoalan sosial yang muncul akibat perkembangan teknologi digital. Namun, ketersediaan instrumen yang secara khusus mengukur kemampuan berpikir kritis siswa pada materi Dampak Sosial Informatika masih terbatas. Penelitian ini bertujuan mengembangkan instrumen kemampuan berpikir kritis siswa kelas VIII SMP pada materi Dampak Sosial Informatika serta memperoleh bukti kelayakannya berdasarkan validitas, reliabilitas, dan kepraktisan. Penelitian menggunakan model pengembangan 4D yang meliputi *define, design, develop, dan disseminate*. Produk yang dikembangkan berupa 18 soal uraian singkat yang mengacu pada enam keterampilan kognitif Facione, yaitu *interpretation, analysis, evaluation, inference, explanation, dan self-regulation*. Validasi isi melibatkan satu akademisi dan dua praktisi. Uji coba terbatas melibatkan 20 siswa kelas VIII SMP Negeri 2 Langsa untuk memperoleh bukti empiris butir dan konsistensi internal instrumen, sedangkan uji coba lebih luas melibatkan 60 siswa. Validitas isi dianalisis menggunakan indeks Aiken's V, validitas empiris menggunakan korelasi Product Moment, dan reliabilitas menggunakan Cronbach's Alpha. Kepraktisan dianalisis berdasarkan respons guru dan siswa. Hasil penelitian menunjukkan nilai rata-rata Aiken's V sebesar 0,909. Pada uji empiris, setelah revisi terhadap butir 4 dan 9, seluruh 18 butir memenuhi kriteria validitas dengan koefisien korelasi 0,735–0,932. Reliabilitas instrumen setelah revisi memperoleh Cronbach's Alpha sebesar 0,975. Respons guru terhadap kepraktisan mencapai 93,75%, sedangkan respons siswa mencapai 90,57%. Hasil tersebut memberikan bukti awal bahwa instrumen memiliki validitas isi, validitas empiris butir, konsistensi internal, dan penerimaan pengguna yang baik. Penelitian lebih lanjut diperlukan untuk menguji validitas konstruk, reliabilitas antarpenilai, karakteristik butir, dan generalisasi instrumen pada sampel yang lebih beragam.

Kata kunci: Instrumen Berpikir Kritis; Dampak Sosial Informatika

Abstract

Critical thinking is an essential competency in Informatics education, particularly when students are confronted with various social issues arising from the development of digital technology. However, the availability of instruments specifically designed to measure students' critical thinking skills in the topic of Social Impacts of Informatics remains limited. This study aimed to develop an instrument for assessing the critical thinking skills of eighth-grade junior high school students on the topic of Social Impacts of Informatics and to obtain evidence of its quality in terms of validity, reliability, and practicality. The study employed the 4D development model, consisting of define, design, develop, and disseminate stages. The developed product comprised 18 short-answer essay items based on six cognitive skills proposed by Facione: interpretation, analysis, evaluation, inference, explanation, and self-regulation. Content validation involved one academic expert and two practitioners. A limited trial was conducted with 20 eighth-grade students at SMP Negeri 2 Langsa to obtain empirical evidence of item validity and the internal consistency of the instrument, while a larger-scale trial involved 60 students. Content validity was analyzed using Aiken's V index, empirical item validity using the Pearson Product-Moment correlation, and reliability using Cronbach's Alpha. Practicality was assessed based on teacher and student responses. The results showed a mean Aiken's V value of 0.909. In the empirical testing, following revisions to Items 4 and 9, all 18 items met the validity criteria, with correlation coefficients ranging from 0.735 to 0.932. The revised instrument demonstrated a Cronbach's Alpha coefficient of 0.975. Teacher responses indicated a practicality

level of 93.75%, while student responses reached 90.57%. These findings provide initial evidence that the instrument demonstrates satisfactory content validity, empirical item validity, internal consistency, and user acceptability. Further research is needed to examine construct validity, inter-rater reliability, item characteristics, and the generalizability of the instrument across more diverse samples.

Keywords: *Critical Thinking Instrument; Social Impacts of Informatics*

PENDAHULUAN

Perkembangan teknologi digital telah mengubah cara individu memperoleh, memproduksi, dan menyebarkan informasi serta berinteraksi dalam kehidupan sehari-hari. Perubahan tersebut memberikan peluang bagi pendidikan untuk mengembangkan kompetensi peserta didik yang tidak hanya berorientasi pada penguasaan pengetahuan, tetapi juga pada kemampuan menggunakan pengetahuan secara kritis dalam menghadapi berbagai persoalan nyata. Kemampuan berpikir kritis menjadi salah satu kompetensi penting dalam konteks tersebut karena memungkinkan peserta didik memahami informasi, menganalisis hubungan antarinformasi, mengevaluasi bukti dan argumen, menarik kesimpulan, menjelaskan alasan, serta merefleksikan proses berpikirnya (OECD, 2024).

Urgensi pengembangan kemampuan berpikir kritis juga memiliki landasan yuridis dalam sistem pendidikan nasional. Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional menempatkan pengembangan potensi peserta didik agar menjadi manusia yang berilmu, cakap, kreatif, mandiri, dan bertanggung jawab sebagai bagian dari tujuan pendidikan nasional. Selanjutnya, kebijakan standar kompetensi lulusan dan implementasi Kurikulum Merdeka memberikan perhatian terhadap kemampuan bernalar kritis sebagai salah satu kompetensi yang perlu dikembangkan dalam pendidikan. Dengan demikian, pengembangan kemampuan berpikir kritis perlu diikuti dengan sistem asesmen yang mampu memberikan informasi mengenai keterampilan tersebut secara terukur.

Dalam pembelajaran Informatika, kebutuhan tersebut semakin relevan karena Informatika tidak hanya berkaitan dengan aspek teknis komputasi, tetapi juga dengan hubungan antara teknologi, individu, dan masyarakat. Materi Dampak Sosial Informatika secara khusus membahas berbagai konsekuensi penggunaan teknologi dalam kehidupan sosial sehingga menyediakan konteks yang dekat dengan pengalaman peserta didik. Permasalahan seperti penyebaran hoaks, keamanan data pribadi, etika penggunaan media sosial, jejak digital, cyberbullying, dan penggunaan teknologi secara bertanggung jawab menuntut siswa untuk tidak sekadar mengingat konsep, tetapi juga memahami informasi, menganalisis sebab-akibat, mengevaluasi kredibilitas informasi, menarik kesimpulan, memberikan alasan, dan merefleksikan perilakunya. Namun, tuntutan tersebut belum sepenuhnya sejalan dengan praktik asesmen di lapangan. Hasil observasi awal di SMP Negeri 2 Langsa menunjukkan bahwa instrumen evaluasi Informatika yang digunakan masih didominasi oleh soal berbasis hafalan, pilihan ganda, dan pemahaman sederhana. Kondisi tersebut menyebabkan proses berpikir siswa belum tergali secara optimal. Temuan observasi tersebut diperkuat melalui wawancara dengan guru Informatika yang menyampaikan bahwa guru mengalami keterbatasan dalam mengembangkan instrumen yang secara khusus mengukur kemampuan berpikir kritis. Kesulitan terutama berkaitan dengan penyusunan indikator, konstruksi soal, rubrik penskoran, dan pemilihan konteks permasalahan yang relevan dengan kehidupan siswa. Tesis penelitian juga mencatat bahwa guru dalam forum MGMP Informatika menghadapi kebutuhan serupa, khususnya untuk materi nonpemrograman seperti Dampak Sosial Informatika (Guzdial, 2015; Denning & Martell, 2015).

Permasalahan tersebut menunjukkan adanya kesenjangan antara tuntutan pembelajaran Informatika dan ketersediaan instrumen asesmen. Pada satu sisi, siswa dihadapkan pada lingkungan digital yang penuh dengan informasi dan berbagai persoalan sosial-teknologi. Pada sisi lain, instrumen yang digunakan belum secara khusus dirancang

untuk mengungkap kemampuan siswa dalam menghadapi persoalan tersebut melalui proses berpikir kritis. Dengan demikian, pengembangan instrumen yang kontekstual menjadi penting agar asesmen tidak hanya mengukur penguasaan materi, tetapi juga memberikan gambaran mengenai kemampuan siswa dalam menggunakan proses berpikir kritis ketika menghadapi permasalahan sosial yang berkaitan dengan teknologi.

Secara teoretis, kemampuan berpikir kritis dalam penelitian ini merujuk pada enam keterampilan kognitif yang dikemukakan Facione, yaitu *interpretation, analysis, evaluation, inference, explanation, dan self-regulation*. *Interpretation* berkaitan dengan kemampuan memahami dan mengungkapkan makna informasi atau situasi; *analysis* berkaitan dengan identifikasi hubungan antara pernyataan, pertanyaan, konsep, atau alasan; *evaluation* berkaitan dengan penilaian terhadap kredibilitas informasi dan kekuatan alasan; *inference* berkaitan dengan penarikan kesimpulan berdasarkan informasi atau bukti; *explanation* berkaitan dengan kemampuan menyampaikan hasil penalaran beserta alasan yang mendasarinya; sedangkan *self-regulation* berkaitan dengan kemampuan memantau dan merefleksikan proses berpikir atau tindakan sendiri (Facione, 2011; Halpern, 2021).

Enam keterampilan tersebut kemudian diterjemahkan ke dalam instrumen berbentuk uraian singkat. Pemilihan bentuk uraian didasarkan pada kebutuhan untuk memperoleh respons siswa yang memperlihatkan alasan, proses penalaran, kesimpulan, dan strategi yang digunakan. Setiap indikator direpresentasikan oleh tiga butir soal, sehingga keseluruhan instrumen terdiri atas 18 butir. Setiap indikator juga dipetakan pada tingkat kognitif C4, C5, dan C6. Dengan demikian, tingkat kognitif Bloom digunakan untuk menggambarkan kompleksitas tuntutan tugas, sedangkan indikator Facione digunakan sebagai dasar operasionalisasi konstruk kemampuan berpikir kritis (Brookhart, 2010; Nitko & Brookhart, 2022).

Penelitian terdahulu telah mengembangkan berbagai instrumen kemampuan berpikir kritis pada sejumlah bidang studi dan jenjang pendidikan. Namun, masih terdapat ruang pengembangan instrumen yang secara khusus mengintegrasikan enam keterampilan Facione dengan konteks Dampak Sosial Informatika pada siswa SMP. Penelitian sebelumnya juga banyak berfokus pada kemampuan berpikir kritis dalam pembelajaran sains atau pada pengukuran kemampuan secara umum, sedangkan konteks persoalan sosial dan etika penggunaan teknologi dalam kehidupan sehari-hari siswa belum menjadi fokus utama. Kesenjangan tersebut menjadi dasar pengembangan instrumen dalam penelitian ini (Ariyani et al., 2021; Marselina et al., 2021; Maryono, 2022).

Kebaruan penelitian terletak pada pengembangan instrumen kemampuan berpikir kritis yang mengintegrasikan enam keterampilan kognitif Facione dengan konteks Dampak Sosial Informatika untuk siswa kelas VIII SMP. Instrumen tidak hanya memuat indikator berpikir kritis secara teoritis, tetapi menerjemahkannya ke dalam butir uraian kontekstual yang berkaitan dengan persoalan digital yang dekat dengan kehidupan siswa serta dilengkapi dengan pedoman penskoran. Penelitian ini juga menggunakan data kebutuhan lapangan dari guru Informatika sebagai dasar pengembangan produk, sehingga instrumen dikembangkan berdasarkan kebutuhan empiris sekaligus landasan teoritis (Bahri & Supahar, 2019; Pacheco & Iturra, 2023; Payan-Carreira et al., 2022).

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan instrumen kemampuan berpikir kritis siswa pada materi Dampak Sosial Informatika dan memperoleh bukti kelayakan instrumen ditinjau dari validitas, reliabilitas, dan kepraktisan. Bukti kelayakan tersebut diharapkan dapat menjadi dasar awal bagi penggunaan instrumen dalam konteks pembelajaran Informatika sekaligus menjadi pijakan bagi penelitian lanjutan mengenai kualitas instrumen kemampuan berpikir kritis pada konteks sosial-teknologi (Butler, 2024).

METODE

Penelitian ini merupakan penelitian pengembangan (*research and development*) yang bertujuan menghasilkan instrumen kemampuan berpikir kritis siswa pada materi

Dampak Sosial Informatika. Model pengembangan yang digunakan adalah model 4D yang terdiri atas empat tahap, yaitu *define, design, develop,* dan *disseminate*.

Pada tahap *define*, dilakukan analisis kebutuhan, karakteristik peserta didik, karakteristik materi, dan tujuan pengukuran. Analisis kebutuhan dilakukan melalui observasi pembelajaran dan wawancara dengan guru Informatika untuk mengidentifikasi kondisi asesmen yang digunakan serta kebutuhan terhadap instrumen kemampuan berpikir kritis. Tahap *design* dilakukan dengan menetapkan konstruk kemampuan berpikir kritis berdasarkan enam keterampilan Facione, menyusun kisi-kisi, menentukan indikator soal, menetapkan tingkat kognitif C4–C6, menyusun butir soal, serta mengembangkan pedoman penskoran analitik. Tahap *develop* meliputi validasi ahli, revisi instrumen, uji coba terbatas, analisis validitas empiris butir, analisis reliabilitas, dan penyempurnaan produk. Validasi melibatkan tiga validator yang terdiri atas satu akademisi dan dua praktisi. Uji coba terbatas dilakukan terhadap 20 siswa kelas VIII SMP Negeri 2 Langsa. Data dari uji coba terbatas digunakan untuk menganalisis validitas empiris butir dan konsistensi internal instrumen. Setelah revisi, instrumen digunakan pada uji coba lebih luas yang melibatkan 60 siswa untuk memperoleh gambaran kemampuan berpikir kritis siswa. Tahap *disseminate* dilakukan secara terbatas melalui penyampaian atau berbagi produk instrumen kepada guru Informatika sebagai pengguna potensial, termasuk dalam konteks forum MGMP Informatika. Diseminasi terbatas tersebut dimaksudkan untuk memperkenalkan produk, bukan untuk mengklaim generalisasi penggunaan instrumen pada seluruh sekolah.

Subjek penelitian terdiri atas tiga validator, yaitu satu akademisi dan dua praktisi, 20 siswa kelas VIII SMP Negeri 2 Langsa pada uji coba terbatas, serta 60 siswa pada uji coba lebih luas. Validator memberikan penilaian terhadap kualitas isi, konstruksi, bahasa, dan kesesuaian instrumen dengan indikator kemampuan berpikir kritis. Guru dan siswa memberikan respons terhadap kepraktisan instrumen.

Data penelitian dikumpulkan menggunakan lembar observasi, pedoman wawancara, lembar validasi ahli, instrumen tes kemampuan berpikir kritis, serta angket respons guru dan siswa. Instrumen tes terdiri atas 18 soal uraian singkat dengan skor maksimum 4 pada setiap butir sehingga skor maksimum keseluruhan adalah 72 (Sugiono, 2019).

Konstruk kemampuan berpikir kritis dioperasionalkan ke dalam enam indikator Facione. Setiap indikator terdiri atas tiga butir soal yang mewakili tuntutan kognitif C4, C5, dan C6. Dengan demikian, setiap butir memiliki satu indikator berpikir kritis utama sebagai target pengukuran.

Validitas isi dianalisis menggunakan indeks Aiken's V berdasarkan penilaian tiga validator. Validitas empiris butir dianalisis menggunakan korelasi Product Moment antara skor setiap butir dan skor total. Reliabilitas internal instrumen dianalisis menggunakan koefisien Cronbach's Alpha.

Kepraktisan instrumen dianalisis berdasarkan respons guru dan siswa menggunakan skala empat tingkat. Persentase kepraktisan dihitung dengan membandingkan skor yang diperoleh dengan skor maksimum kemudian dikalikan 100%. Kategori kepraktisan menggunakan interval 81–100% sangat praktis, 61–80% praktis, 41–60% cukup praktis, 21–40% kurang praktis, dan ≤20% tidak praktis.

Interpretasi hasil kemampuan berpikir kritis siswa pada uji coba lebih luas dilakukan berdasarkan persentase skor terhadap skor maksimum 72. Kategori kemampuan ditentukan menggunakan interval persentase yang telah ditetapkan dalam penelitian, yaitu 81–100% sangat kritis, 61–80% kritis, 41–60% cukup kritis, 21–40% kurang kritis, dan 0–20% sangat kurang kritis.

Tabel 1. Operasionalisasi Kemampuan Berpikir Kritis

Indikator Facione	Definisi operasional	Konteks/materi	Nomor soal	Level kognitif
Interpretation	Kemampuan memahami, mengidentifikasi, dan mengungkapkan makna informasi atau fenomena sosial yang berkaitan dengan teknologi	Dampak teknologi terhadap interaksi sosial dan media sosial	1–3	C4–C6
Analysis	Kemampuan mengidentifikasi hubungan antara informasi, faktor, alasan, atau sebab-akibat dalam suatu permasalahan	Penggunaan media sosial dan penyebaran hoaks	4–6	C4–C6
Evaluation	Kemampuan menilai kredibilitas informasi, kekuatan alasan, atau kelayakan suatu kebijakan berdasarkan kriteria dan alasan yang relevan	Kredibilitas informasi dan kebijakan penggunaan internet	7–9	C4–C6
Inference	Kemampuan menarik kesimpulan atau membuat prediksi berdasarkan informasi dan bukti yang tersedia	Permainan daring, kebocoran data pribadi, dan literasi digital	10–12	C4–C6
Explanation	Kemampuan menyampaikan hasil penalaran, alasan, dan solusi secara logis dan sistematis	Etika media sosial dan penyebaran hoaks	13–15	C4–C6
Self-regulation	Kemampuan memantau, mengevaluasi, dan merencanakan perbaikan terhadap proses berpikir atau perilaku penggunaan teknologi	Verifikasi informasi dan penggunaan teknologi secara bertanggung jawab	16–18	C4–C6

HASIL DAN PEMBAHASAN

Pengembangan Instrumen

Produk akhir penelitian berupa instrumen kemampuan berpikir kritis siswa pada materi Dampak Sosial Informatika dalam bentuk 18 soal uraian singkat. Instrumen disusun berdasarkan enam indikator Facione, yaitu *interpretation, analysis, evaluation, inference, explanation, dan self-regulation*. Setiap indikator direpresentasikan oleh tiga butir soal

sehingga keseluruhan instrumen terdiri atas 18 butir. Setiap butir diberi skor maksimum 4 sehingga skor maksimum keseluruhan adalah 72 (Facione, 2011; Brookhart, 2010).

Pengembangan dilakukan melalui model 4D yang meliputi tahap *define, design, develop,* dan *disseminate*. Pada tahap *define*, dilakukan identifikasi kebutuhan melalui observasi dan wawancara untuk mengetahui kondisi pembelajaran dan asesmen Informatika. Tahap *design* menghasilkan kisi-kisi berdasarkan enam indikator Facione dan tingkat kognitif C4–C6. Tahap *develop* mencakup validasi ahli, revisi, uji coba terbatas, analisis empiris, dan penyempurnaan instrumen. Selanjutnya, tahap *disseminate* dilakukan secara terbatas melalui penyampaian produk kepada pengguna potensial, khususnya guru Informatika.

Konteks soal dikembangkan berdasarkan permasalahan yang dekat dengan kehidupan digital siswa, meliputi penyebaran hoaks, keamanan data pribadi, etika digital, jejak digital, *cyberbullying*, penggunaan media sosial, dan literasi digital. Dengan demikian, stimulus soal tidak hanya menguji penguasaan materi, tetapi memberikan situasi yang memungkinkan siswa menunjukkan proses berpikir kritis (McGrew et al., 2018; OECD, 2024).

Validitas Isi

Validitas isi diperoleh melalui penilaian tiga validator yang terdiri atas satu akademisi dan dua praktisi. Penilaian menggunakan skala 1–5 dan dianalisis menggunakan indeks Aiken's V (Aiken, 1985; Hidayah & Muhtarom, 2023; Suhardi, 2022). Hasil validasi disajikan pada Tabel 2.

Tabel 2. Hasil Validitas Isi Berdasarkan Aiken's V

No.	Indikator Penilaian	Aiken's V
1	Instrumen selaras dengan tujuan pembelajaran	0,917
2	Aktivitas asesmen mendukung pembelajaran bermakna	0,917
3	Instrumen mendukung kemampuan berpikir kritis	1,000
4	Struktur instrumen sistematis dan logis	0,917
5	Instruksi tugas jelas dan operasional	0,833
6	Instrumen mengukur berpikir kritis secara tepat	0,917
7	Hasil asesmen dapat menjadi dasar umpan balik	0,833
8	Soal mencerminkan indikator berpikir kritis	1,000
9	Tingkat kognitif mengarah HOTS	1,000
10	Tingkat kesulitan sesuai karakteristik siswa	0,833
11	Instrumen layak digunakan sebagai asesmen berpikir kritis	0,833
	Rata-rata	0,909

Sumber: data penelitian.

Nilai Aiken's V berada pada rentang 0,833–1,000 dengan rata-rata 0,909. Seluruh indikator berada di atas batas 0,80 yang digunakan dalam penelitian. Hasil ini menunjukkan adanya kesepakatan yang tinggi dari validator terhadap kesesuaian instrumen dengan tujuan pembelajaran, konstruk kemampuan berpikir kritis, materi, konstruksi soal, dan karakteristik peserta didik. Data tersebut juga menunjukkan bahwa aspek "instrumen mendukung kemampuan berpikir kritis", "soal mencerminkan indikator berpikir kritis", dan "tingkat kognitif mengarah HOTS" memperoleh nilai tertinggi sebesar 1,000. Namun, nilai tersebut merupakan hasil Aiken's V berdasarkan 11 indikator penilaian instrumen, bukan nilai Aiken's V untuk setiap dari 18 butir soal. Oleh karena itu, hasil ini lebih tepat disebut sebagai bukti validitas isi pada tingkat aspek/indikator penilaian. Analisis Aiken's V per butir dapat dilakukan apabila data mentah penilaian masing-masing validator tersedia (Aiken, 1985; Kurniawan et al., 2025; Noviana et al., 2023).

Validitas Empiris Butir

Setelah proses validasi ahli dan revisi, instrumen diuji secara empiris. Uji coba terbatas melibatkan 20 siswa kelas VIII SMP Negeri 2 Langsa. Validitas empiris dianalisis dengan korelasi antara skor butir dan skor total (Azwar, 2015; Audiwinanda & Mahmudi, 2024).

Pada pengujian awal, terdapat dua butir yang memerlukan perhatian, yaitu butir 4 dan 9. Kedua butir tersebut kemudian diperbaiki berdasarkan hasil analisis dan pertimbangan terhadap redaksi serta kesesuaian tuntutan soal. Setelah revisi, seluruh butir menunjukkan koefisien korelasi yang positif.

Tabel 3. Validitas Empiris Butir Sebelum dan Sesudah Revisi

Butir	Korelasi sebelum revisi*	Status	Korelasi setelah revisi	Status
1	0,881	Valid	0,932	Valid
2	0,842	Valid	0,826	Valid
3	0,781	Valid	0,769	Valid
4	0,512	Perlu revisi	0,813	Valid
5	0,812	Valid	0,827	Valid
6	0,798	Valid	0,792	Valid
7	0,741	Valid	0,735	Valid
8	0,833	Valid	0,845	Valid
9	0,421	Perlu revisi	0,790	Valid
10	0,763	Valid	0,801	Valid
11	0,732	Valid	0,760	Valid
12	0,835	Valid	0,846	Valid
13	0,762	Valid	0,792	Valid
14	0,832	Valid	0,845	Valid
15	0,711	Valid	0,752	Valid
16	0,789	Valid	0,818	Valid
17	0,748	Valid	0,792	Valid
18	0,739	Valid	0,769	Valid

*Berdasarkan *corrected item-total correlation* pada output analisis sebelum revisi.
Sumber: data penelitian.

Output SPSS menunjukkan bahwa setelah revisi, *corrected item-total correlation* seluruh butir berada pada rentang 0,735–0,932. Nilai tertinggi diperoleh butir 1 sebesar 0,932, sedangkan nilai terendah diperoleh butir 7 sebesar 0,735. Seluruh koefisien menunjukkan hubungan positif antara skor butir dan skor total.

Perubahan paling terlihat terjadi pada butir 4 dan 9. Koefisien butir 4 meningkat dari 0,512 menjadi 0,813, sedangkan butir 9 meningkat dari 0,421 menjadi 0,790. Dengan demikian, proses revisi memberikan perbaikan terhadap keterkaitan kedua butir dengan skor total. Hasil ini memperlihatkan bahwa analisis empiris dapat digunakan sebagai dasar untuk mengidentifikasi butir yang memerlukan penyempurnaan sebelum instrumen digunakan pada tahap berikutnya.

Perlu ditegaskan bahwa validitas empiris tersebut merupakan bukti hubungan internal butir dengan skor total pada sampel penelitian. Hasil tersebut belum dapat ditafsirkan sebagai bukti validitas konstruk secara komprehensif karena penelitian ini belum melakukan analisis faktor atau pengujian struktur konstruk.

Reliabilitas Instrumen

Reliabilitas instrumen dianalisis menggunakan Cronbach’s Alpha terhadap 18 butir pada data uji coba terbatas yang melibatkan 20 siswa. Hasil analisis menunjukkan peningkatan koefisien reliabilitas setelah revisi (Cronbach, 1951).

Tabel 4. Reliabilitas Instrumen

Tahap Pengujian	Cronbach's Alpha	Jumlah Butir	Keterangan
Sebelum revisi	0,958	18	Sangat tinggi
Setelah revisi	0,975	18	Sangat tinggi

Sumber: data penelitian.

Nilai Cronbach's Alpha meningkat dari 0,958 sebelum revisi menjadi 0,975 setelah revisi. Output SPSS setelah revisi juga menunjukkan bahwa analisis dilakukan terhadap 18 butir dengan nilai Alpha 0,975. Nilai tersebut menunjukkan konsistensi internal yang tinggi pada data uji coba penelitian. Dengan demikian, skor yang dihasilkan oleh 18 butir memiliki konsistensi internal yang baik dalam konteks sampel penelitian. Meskipun demikian, reliabilitas Cronbach's Alpha tidak mencakup reliabilitas antarpenilai. Karena instrumen berbentuk uraian dan setiap jawaban diberi skor menggunakan rubrik, pengujian *inter-rater reliability* masih diperlukan untuk memperoleh bukti yang lebih lengkap mengenai konsistensi penskoran.

Hasil Uji Coba Lebih Luas

Setelah melalui proses validasi, revisi, dan uji coba terbatas, instrumen digunakan pada 60 siswa untuk memperoleh gambaran kemampuan berpikir kritis siswa pada materi Dampak Sosial Informatika. Instrumen terdiri atas 18 butir dengan skor maksimum 72.

Hasil pengukuran menunjukkan skor tertinggi 69 dan skor terendah 18. Rata-rata skor siswa adalah 42,18 dari skor maksimum 72 atau sebesar 58,59%. Berdasarkan interval kategori yang digunakan dalam penelitian, rata-rata tersebut termasuk kategori cukup kritis.

Tabel 5. Distribusi Kemampuan Berpikir Kritis Siswa

Kategori	Rentang (%)	Jumlah	Persentase (%)
Sangat kritis	81–100	4	6,67
Kritis	61–80	12	20,00
Cukup kritis	41–60	25	41,67
Kurang kritis	21–40	15	25,00
Sangat kurang kritis	0–20	4	6,67
Total		60	100,00

Sumber: data penelitian.

Sebagian besar siswa berada pada kategori cukup kritis, yaitu 25 siswa (41,67%). Sebanyak 12 siswa (20,00%) berada pada kategori kritis, sedangkan 15 siswa (25,00%) berada pada kategori kurang kritis. Empat siswa (6,67%) termasuk sangat kritis dan empat siswa lainnya (6,67%) termasuk sangat kurang kritis. Distribusi tersebut menunjukkan adanya variasi kemampuan berpikir kritis siswa.

Hasil tersebut tidak digunakan untuk menyatakan bahwa instrumen meningkatkan kemampuan berpikir kritis siswa karena penelitian tidak menggunakan desain eksperimen atau pengukuran sebelum dan sesudah perlakuan. Data 60 siswa dalam penelitian ini lebih tepat digunakan untuk menggambarkan distribusi kemampuan siswa setelah instrumen dikembangkan.

Kemampuan Berdasarkan Indikator Facione

Analisis berdasarkan enam indikator menunjukkan bahwa kemampuan siswa berbeda pada setiap aspek.

Tabel 6. Kemampuan Berpikir Kritis Berdasarkan Indikator

Indikator	Rata-rata Skor	Persentase	Kategori
Interpretation	7,4	61,67%	Kritis
Analysis	6,9	57,50%	Cukup kritis
Evaluation	6,5	54,17%	Cukup kritis

Indikator	Rata-rata Skor	Persentase	Kategori
Inference	6,3	52,50%	Cukup kritis
Explanation	6,1	50,83%	Cukup kritis
Self-regulation	5,8	48,33%	Cukup kritis

Sumber: data penelitian.

Interpretation memperoleh persentase tertinggi, yaitu 61,67%, dan termasuk kategori kritis. Hasil tersebut menunjukkan bahwa siswa relatif lebih mampu memahami makna informasi dan mengidentifikasi persoalan dalam fenomena Dampak Sosial Informatika.

Analysis memperoleh 57,50%. Hal ini menunjukkan bahwa kemampuan siswa dalam menghubungkan berbagai informasi dan mengidentifikasi hubungan sebab-akibat masih berada pada kategori cukup kritis. *Evaluation* memperoleh 54,17%, menunjukkan bahwa kemampuan dalam menilai kredibilitas informasi dan kekuatan argumen masih perlu diperkuat.

Inference memperoleh 52,50%, sedangkan *explanation* memperoleh 50,83%. Kedua hasil tersebut menunjukkan bahwa siswa masih menghadapi kesulitan ketika harus menarik kesimpulan berdasarkan bukti serta menjelaskan alasan secara logis dan sistematis. *Self-regulation* memperoleh nilai terendah, yaitu 48,33%. Hasil ini menunjukkan bahwa tugas yang meminta siswa merefleksikan atau mengevaluasi perilaku penggunaan teknologi dan menyusun strategi perbaikan masih relatif lebih menantang.

Kepraktisan Berdasarkan Respons Guru

Kepraktisan instrumen dinilai oleh satu guru Informatika melalui angket yang terdiri atas 20 pernyataan menggunakan skala empat tingkat. Aspek yang dinilai meliputi kesesuaian isi dan materi, kebahasaan dan penyajian, kepraktisan, serta manfaat dan kelayakan.

Tabel 7. Respons Guru terhadap Instrumen

Aspek	Jumlah Butir	Skor Perolehan	Skor Maksimum	Persentase
Kesesuaian isi dan materi	9	34	36	94,44%
Kebahasaan dan penyajian	4	15	16	93,75%
Kepraktisan	3	11	12	91,67%
Manfaat dan kelayakan	4	15	16	93,75%
Total	20	75	80	93,75%

Sumber: data penelitian.

Respons guru secara keseluruhan mencapai 93,75%. Nilai tersebut berada pada kategori sangat baik berdasarkan kriteria yang digunakan dalam penelitian. Kesesuaian isi dan materi memperoleh 94,44%, kebahasaan dan penyajian 93,75%, kepraktisan 91,67%, serta manfaat dan kelayakan 93,75%. Hasil tersebut menunjukkan bahwa dari perspektif satu guru pengguna, instrumen memiliki karakteristik yang mudah diterapkan dan sesuai dengan kebutuhan asesmen. Namun, karena hanya melibatkan satu guru, hasil kepraktisan tersebut tidak dapat dianggap sebagai representasi seluruh guru Informatika.

Kepraktisan Berdasarkan Respons Siswa

Respons siswa diperoleh dari 20 siswa yang mengikuti uji coba terbatas. Angket terdiri atas 11 pernyataan dengan empat aspek, yaitu kemudahan penggunaan, pemahaman materi, dukungan terhadap berpikir kritis, dan aspek motivasional.

Tabel 8. Respons Siswa terhadap Instrumen

Aspek	Jumlah Butir	Skor Perolehan	Skor Maksimum	Persentase
Kemudahan penggunaan	5	375	400	93,75%
Pemahaman materi	2	144	160	90,00%
Dukungan terhadap berpikir kritis	2	140	160	87,50%
Aspek motivasional	2	138	160	86,25%
Total	11	797	880	90,57%

Sumber: data penelitian.

Persentase keseluruhan respons siswa sebesar 90,57% berada pada kategori sangat baik. Aspek kemudahan penggunaan memperoleh nilai tertinggi sebesar 93,75%, diikuti pemahaman materi sebesar 90,00%, dukungan terhadap berpikir kritis sebesar 87,50%, dan aspek motivasional sebesar 86,25%.

Hasil tersebut menunjukkan bahwa siswa memberikan respons positif terhadap penggunaan instrumen. Namun, persentase pada aspek motivasional hanya menunjukkan persepsi siswa terhadap pengalaman menggunakan instrumen. Oleh sebab itu, temuan tersebut tidak dapat digunakan untuk menyimpulkan bahwa instrumen meningkatkan motivasi belajar siswa.

Kualitas Pengembangan Instrumen

Hasil pengembangan menunjukkan bahwa instrumen memiliki struktur yang jelas antara konstruk, indikator, konteks materi, bentuk soal, dan pedoman penskoran. Penggunaan enam indikator Facione memungkinkan kemampuan berpikir kritis diterjemahkan menjadi perilaku kognitif yang dapat diamati melalui respons siswa. Pembagian tiga butir untuk setiap indikator dilakukan untuk memperoleh representasi yang seimbang. Setiap indikator dipetakan ke dalam tiga tingkat tuntutan kognitif, yaitu C4, C5, dan C6. Dalam struktur ini, Facione digunakan sebagai kerangka konstruk berpikir kritis, sedangkan Taksonomi Bloom digunakan untuk menggambarkan kompleksitas tuntutan kognitif. Dengan demikian, kedua kerangka tersebut memiliki fungsi yang berbeda dan tidak diposisikan sebagai konstruk yang sama (Facione, 2011; Brookhart, 2010).

Konteks Dampak Sosial Informatika memberikan karakteristik khusus terhadap instrumen. Permasalahan seperti hoaks, keamanan data pribadi, etika digital, jejak digital, *cyberbullying*, dan media sosial memungkinkan siswa memberikan respons berdasarkan situasi yang dekat dengan kehidupan mereka. Konteks tersebut juga membuat pengukuran tidak hanya berorientasi pada penguasaan istilah atau konsep, tetapi pada penerapan proses berpikir ketika menghadapi masalah sosial yang berkaitan dengan teknologi (McGrew et al., 2018; Guzdial, 2015).

Validitas Isi dan Revisi Butir

Nilai rata-rata Aiken's V sebesar 0,909 memberikan bukti bahwa validator menilai instrumen memiliki kesesuaian isi yang tinggi. Nilai terendah 0,833 tetap berada di atas kriteria yang ditetapkan dalam penelitian, sedangkan tiga indikator memperoleh nilai maksimum 1,000. Meskipun demikian, hasil ini perlu ditafsirkan secara proporsional. Aiken's V yang diperoleh merupakan agregasi berdasarkan 11 indikator penilaian. Oleh karena itu, hasil tersebut menunjukkan dukungan terhadap validitas isi instrumen pada tingkat aspek penilaian, tetapi belum memberikan informasi Aiken's V untuk setiap butir secara individual (Aiken, 1985; Hidayah & Muhtarom, 2023; Noviana et al., 2023).

Revisi butir 4 dan 9 menjadi bagian penting dalam proses pengembangan. Setelah revisi, kedua butir mengalami peningkatan *corrected item-total correlation*, masing-masing menjadi 0,813 dan 0,790. Secara keseluruhan, output setelah revisi menunjukkan seluruh 18 butir mempunyai korelasi positif dengan skor total, dengan nilai 0,735–0,932.

Temuan tersebut memperlihatkan pentingnya menggabungkan pertimbangan teoritis dan empiris dalam pengembangan instrumen. Butir yang secara konseptual telah disusun berdasarkan indikator tetap perlu diperiksa melalui data empiris agar kelemahan redaksi atau karakteristik butir dapat diketahui sebelum instrumen digunakan lebih lanjut (Nieveen, 2021).

Reliabilitas dan Penskoran Soal Uraian

Nilai Cronbach's Alpha sebesar 0,975 menunjukkan konsistensi internal yang sangat tinggi pada 18 butir dalam uji coba terbatas. Karena instrumen menggunakan soal uraian, skor setiap respons siswa ditentukan menggunakan pedoman penskoran dengan rentang skor 1–4. Dengan demikian, reliabilitas yang diperoleh merupakan konsistensi internal skor yang dihasilkan dari respons siswa. Namun, tingginya Alpha tidak otomatis menunjukkan bahwa proses pemberian skor bebas dari subjektivitas (Cronbach, 1951; Azwar, 2015).

Hal tersebut menjadi alasan penting untuk membedakan antara *internal consistency* dan *inter-rater reliability*. Penelitian ini telah memberikan bukti mengenai konsistensi internal, tetapi belum menguji kesepakatan antarpenilai. Penelitian selanjutnya perlu menggunakan minimal dua penilai dan menghitung koefisien kesepakatan yang sesuai agar kualitas penskoran uraian dapat dibuktikan secara lebih komprehensif.

Profil Kemampuan Berpikir Kritis Siswa

Rata-rata kemampuan berpikir kritis siswa sebesar 58,59% berada pada kategori cukup kritis. Distribusi menunjukkan bahwa kelompok terbesar berada pada kategori cukup kritis, yaitu 41,67%, sementara 25% berada pada kategori kurang kritis.

Temuan tersebut memberikan informasi diagnostik mengenai kemampuan siswa, tetapi tidak dapat ditafsirkan sebagai dampak penggunaan instrumen. Instrumen dalam penelitian ini berfungsi untuk mengukur dan menggambarkan kemampuan siswa, bukan sebagai perlakuan pembelajaran yang dirancang untuk meningkatkan kemampuan berpikir kritis.

Jika ditinjau berdasarkan indikator, *interpretation* merupakan aspek dengan capaian tertinggi, sedangkan self-regulation merupakan aspek dengan capaian terendah. Perbedaan tersebut dapat dipahami dari karakteristik tuntutan setiap indikator. Memahami makna informasi relatif lebih dekat dengan pengalaman siswa dibandingkan tugas yang menuntut refleksi terhadap proses berpikir dan perilaku penggunaan teknologi sendiri. Hasil tersebut memperlihatkan bahwa instrumen dapat digunakan sebagai salah satu sumber informasi diagnostik bagi guru mengenai aspek kemampuan berpikir kritis yang relatif lebih kuat maupun yang masih memerlukan perhatian. Namun, interpretasi tersebut tetap terbatas pada konteks sampel dan materi yang digunakan dalam penelitian (Facione, 2011; Butler, 2024).

Kepraktisan Instrumen

Respons guru sebesar 93,75% dan respons siswa sebesar 90,57% menunjukkan penerimaan pengguna yang sangat baik terhadap instrumen. Pada guru, aspek kesesuaian isi dan materi memperoleh 94,44%, sedangkan pada siswa aspek kemudahan penggunaan memperoleh nilai tertinggi 93,75%.

Hasil tersebut memperlihatkan bahwa instrumen tidak hanya memiliki dukungan dari sisi validitas isi dan konsistensi internal, tetapi juga memperoleh respons positif ketika digunakan oleh guru dan siswa. Kepraktisan menjadi aspek penting dalam pengembangan instrumen karena instrumen yang secara teoritis baik tetap perlu dapat diterapkan secara realistis dalam konteks pembelajaran. Akan tetapi, jumlah guru yang memberikan respons hanya satu orang. Oleh karena itu, hasil tersebut lebih tepat dipahami sebagai bukti awal penerimaan pengguna, bukan sebagai bukti kepraktisan yang berlaku secara umum. Pengujian pada lebih banyak guru dan sekolah diperlukan untuk memperoleh gambaran yang lebih representatif (Nieveen, 2021).

Posisi Penelitian terhadap Penelitian Sebelumnya

Penelitian ini memiliki posisi yang berbeda dari pengembangan instrumen berpikir kritis yang menggunakan konteks pembelajaran umum atau bidang studi lain. Fokus penelitian diarahkan secara spesifik pada siswa kelas VIII SMP dan materi Dampak Sosial Informatika dengan konteks persoalan digital yang dekat dengan kehidupan siswa.

Kontribusi utama penelitian bukan semata-mata pada penggunaan teori Facione, tetapi pada penerjemahan enam keterampilan kognitif tersebut ke dalam 18 soal uraian yang terstruktur dan kontekstual. Integrasi antara indikator Facione, materi Dampak Sosial Informatika, tingkat kognitif C4–C6, serta pedoman penskoran memberikan karakteristik khusus pada produk yang dikembangkan.

Dengan demikian, penelitian ini mengisi ruang pengembangan instrumen pada konteks Informatika, khususnya materi yang berkaitan dengan konsekuensi sosial penggunaan teknologi. Instrumen dapat menjadi alternatif sumber asesmen bagi guru, tetapi penggunaannya tetap perlu mempertimbangkan keterbatasan bukti psikometrik yang diperoleh dalam penelitian ini (Bahri & Supahar, 2019; Butler, 2024).

PENUTUP

Penelitian ini menghasilkan instrumen kemampuan berpikir kritis pada materi Dampak Sosial Informatika yang terdiri atas 18 soal uraian singkat berdasarkan enam indikator Facione, yaitu *interpretation, analysis, evaluation, inference, explanation, dan self-regulation*. Setiap indikator direpresentasikan oleh tiga butir yang dikembangkan pada tingkat kognitif C4, C5, dan C6. Hasil validasi isi menunjukkan rata-rata Aiken's V sebesar 0,909 dengan rentang 0,833–1,000. Setelah revisi, seluruh 18 butir menunjukkan *corrected item-total correlation* pada rentang 0,735–0,932 dengan r-tabel 0,444. Reliabilitas instrumen menunjukkan Cronbach's Alpha sebesar 0,975. Respons pengguna menunjukkan kepraktisan sebesar 93,75% dari guru dan 90,57% dari siswa. Berdasarkan hasil tersebut, instrumen menunjukkan bukti awal validitas isi dan validitas empiris butir, konsistensi internal yang sangat tinggi, serta penerimaan pengguna yang baik dalam konteks penelitian. Hasil tersebut belum dimaksudkan sebagai bukti validitas konstruk secara komprehensif atau dasar generalisasi penggunaan instrumen pada seluruh siswa SMP.

Penelitian selanjutnya disarankan melakukan pengujian pada jumlah responden yang lebih besar dan melibatkan sekolah dengan karakteristik yang beragam. Selain itu, karena instrumen berbentuk uraian, penelitian selanjutnya perlu menguji *inter-rater reliability*. Analisis menggunakan Item Response Theory juga dapat digunakan untuk memperoleh informasi yang lebih mendalam mengenai karakteristik setiap butir.

DAFTAR RUJUKAN

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Ariyani, N. P. A. T., Suastra, I. W., & Adnyana, P. B. (2021). Pengembangan instrumen kemampuan berpikir kritis dan keterampilan proses sains siswa kelas V SD. *Jurnal Penelitian dan Evaluasi Pendidikan Indonesia, 11*(2), 69–80. https://doi.org/10.23887/jpepi.v11i2.616

Audiwinanda, S., & Mahmudi, A. (2025). Validity and reliability of students' mathematical communication instruments in class VII middle school statistics material. *Jurnal Penelitian dan Evaluasi Pendidikan, 29*(1), 1–12. https://doi.org/10.21831/pep.v29i1.71234

Azwar, S. (2015). *Reliabilitas dan validitas*. Pustaka Pelajar.

Bahri, M. F., & Supahar. (2019). Content validity and reliability analysis of integrated Islamic-science test instrument to measure student critical thinking ability. *Islam Realitas:*

Journal of Islamic & Social Studies, 5(1), 42–51. https://doi.org/10.30983/islam_realitas.v5i1.894

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.

Butler, H. A. (2024). Predicting everyday critical thinking: A review of critical thinking assessments. *Journal of Intelligence, 12*(2), Article 16. https://doi.org/10.3390/jintelligence12020016

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Denning, P. J., & Martell, C. H. (2015). *Great principles of computing*. Princeton University Press.

Facione, P. A. (2011). *Critical thinking: What it is and why it counts*. Insight Assessment.

Guzdial, M. (2015). *Learner-centered design of computing education: Research on computing for everyone*. Morgan & Claypool Publishers.

Halpern, D. F. (2021). *Thought and knowledge: An introduction to critical thinking* (6th ed.). Routledge. https://doi.org/10.4324/9781003025412

Hidayah, N., & Muhtarom. (2023). Validity and reliability test of teaching materials using Aiken's V formula and SPSS 22. *Schola, 1*(2), 75–82. https://doi.org/10.26877/schola.v1i2.342

Kurniawan, K., Parlindungan, S., & Duden, S. (2025). Aiken's V analysis and Rasch modeling to determine the quality of physics creative thinking test instruments. *Jurnal Riset dan Kajian Pendidikan Fisika, 12*(1), 32–40. https://doi.org/10.12928/jrkpf.v12i1.1161

Marselina, [inisial dan nama penulis perlu diverifikasi], Lasmawan, [inisial perlu diverifikasi], & Dantes, N. (2021). Pengembangan instrumen kemampuan berpikir kritis dan hasil belajar IPS pada siswa kelas V SD. *Jurnal Penelitian dan Evaluasi Pendidikan Indonesia, 11*(2), 105–115. https://doi.org/10.23887/jpepi.v11i2.620

Maryono. (2022). Uji validasi dan reliabilitas instrumen penilaian kemampuan berpikir kritis siswa kelas VI sekolah dasar. *Jurnal Kualita Pendidikan, 3*(3), 146–151.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education, 46*(2), 165–193. https://doi.org/10.1080/00933104.2017.1416320

Nieveen, N. (2021). Educational design research: Quality criteria for educational products. In T. Plomp & N. Nieveen (Eds.), *Educational design research*. Netherlands Institute for Curriculum Development.

Nitko, A. J., & Brookhart, S. M. (2022). *Educational assessment of students* (8th ed.). Pearson.

Noviana, E., Sari, I. K., Zufriady, Erlinda, S., Syafriafdi, N., & Suarmika, P. E. (2023). Instrument validity for learning resources analysis of digital disaster for coastal areas: An analysis of Aiken's V. *Primary: Jurnal Pendidikan Guru Sekolah Dasar, 12*(1), 114–125. https://doi.org/10.33578/jpfkip.v12i1.9320

OECD. (2024). *PISA 2022 results (Volume III): Creative minds, creative schools*. OECD Publishing. https://doi.org/10.1787/765ee8c2-en

Pacheco, C. S., & Iturra, C. (2023). Development and validation of the Complex Thinking Assessment Instrument (CTAI). *Thinking Skills and Creativity, 48*, Article 101295. https://doi.org/10.1016/j.tsc.2023.101305

Payan-Carreira, R., Sacau-Fontenla, A., Rebelo, H., Sebastião, L., & Pnevmatikos, D. (2022). Development and validation of a critical thinking assessment-scale short form. *Education Sciences, 12*(12), Article 938. https://doi.org/10.3390/educsci12120938

Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Suhardi, I. (2022). Perangkat instrumen pengembangan paket soal jenis pilihan ganda menggunakan pengukuran validitas konten formula Aiken's V. *Jurnal Pendidikan Tambusai, 6*(1), 4158–4170. https://jptam.org/index.php/jptam/article/view/3519